大数据挖掘与分析
ISSN 2096-0654 11/25 pp718－729 DOI: 10.26599/
BDMA.2023.9020035
第 7 卷,第 3 期,2024 年 9 月

# 变分自动编码器的评估
# 信用卡异常检测

Faleh Alshameri* 和夏冉

摘要:异常检测是网络安全领域中众多具有挑战性的领域之一。异常可能发生在
多种形式,例如欺诈性信用卡交易、网络入侵和异常图像或
文档。异常检测中最常见的挑战之一是正常状态的模糊性和
缺乏异常样本。传统上,这个问题是通过使用重采样技术或
选择接近正常状态分布的模型。变分自动编码器 (VAE) 具有
尽管更适合生成任务,但在异常检测中尚未得到研究。本研究旨在探索
VAE 在信用卡异常检测中的应用,并评估潜在空间采样技术。在此
研究中,我们评估了基于卷积网络的 VAE 模型在信用卡交易中的使用情况
数据集。我们训练两个 VAE 模型,一个模型包含大量正常数据,另一个模型包含少量
异常数据。我们比较了两种 VAE 模型的性能,并评估了两种 VAE 的潜在空间
通过使用重构误差向量重新缩放模型,我们能够提高模型的预测准确率。我们还比较了 VAE 的有效性
当在不平衡数据集上训练时,模型与其他异常检测模型一样。

关键词:异常检测;优化;不平衡数据集;生成模型;卷积神经网络
网络（CNN）;变分自编码器（VAE）;潜在空间缩放;重构误差

## 1 简介

随着互联网的爆炸式发展[1],针对网络和计算机系统的网络攻击也迅速增加。

私营企业和政府机构对安全在线交易的需求不断增长[2] 。然而,
欺诈检测对于大多数电子商务组织来说仍然具有挑战性。

异常检测是机器学习的一个主题,它受到了很多关注,因为它对许多
应用,包括网络分析、入侵检测、欺诈检测、恶意软件检测、健康监测、脑部扫描、视
频中的异常检测和物联网 (IoT) [3-5] 。

异常检测的重要性在于,在广泛的应用领域中,数据中的异常可以产生大量甚
至关键的信息[3]

。

本研究采用一种算法方法,专注于对持卡人的消费行为进行建模,并将异常数
据视为统计异常值,与正常消费正常值进行比较[6]

。

变分自编码器（VAE)是一种深度学习架构,它是一种特殊类型的自编码器,由
Kingma 和 Welling[7]以及 Islam 等人[8]首次提出。

。VAE 属于概率图模型 (PGM) 家族。VAE 被定义为有向概
率图模型,它是通过近似人工

● Faleh Alshameri 就职于马里兰大学全球校区商学院,地址:美国马里兰州
阿德尔菲 20783。电子邮箱:faleh.alshameri@faculty.umgc.edu。

● 夏冉就职于美国弗吉尼亚州阿灵顿市玛丽蒙特大学技术与创新学院、商学
院、创新、领导力与技术学院,邮编 22207。电子邮箱:
r0x28181@marymount.edu。

*通信地址。
收稿日期:2023-07-16;修订日期:2023-11-15;
接受日期:2023-11-22

神经网络的后验[1] 。

尽管 VAE 是作为输入重构的生成模型而发明的,但最近的研究已经表明,通过在原始架构上进行各种修改,VAE 在异常检测方面具有强大的潜力。在数据生成中,VAE 由编码器-解码器结构组成,其中编码器将输入数据转换为低维潜在表示,解码器将此潜在表示转换回与输入数据相同的维度[7]

。 这两个模型相辅相成。为了使生成模型在期望最大化学习的迭代过程中更新其参数,编码器或识别模型为其后验潜在随机变量提供近似值。识别模型使用解码器或生成模型作为支架来学习数据的准确表示,其中可能包括类标签。根据贝叶斯规则,识别模型大致对应于生成模型的逆[7] 。

使用 VAE 进行异常检测的直觉来自于它如何近似观察到的数据采样的真实分布。这假设在异常检测中,正常数据和异常数据是从不同的分布生成的。一旦 VAE 模型在正常数据上得到充分训练,它就会学习其分布并创建一个连续的潜在

低维表示空间。因此,如果将异常数据作为输入传递给此 VAE 模型,则重构误差自然会高于将正常数据作为输入。此外,潜在空间中的异常输入应形成一些与正常输入不同的簇形状。

本研究论文的目的是评估卷积 VAE 在信用卡欺诈检测数据集上的使用情况。评估包括分别在正常数据和异常数据上训练两个 VAE 模型。然后使用混淆矩阵和 F1 分数比较两个模型的性能。我们还通过检查异常数据在潜在空间中的表示和聚类方式以及潜在空间缩放对模型性能的影响来评估潜在空间。模型将数据集划分为目标的能力由 VAE 的潜在空间表示。通过正确读取潜在空间边界可以生成新的数据样本[8]

。

本文的其余部分组织如下:

第 2 部分 :文献综述部分,探讨 VAE 和异常检测的相关工作。第 3 部分 :研究方法部分,解释数据来源、探索性数据分析和模型构建。第 4 部分 :数据分析和结果部分,评估模型性能以及潜在空间缩放。最后,第 5 部分 :讨论和结论部分,总结本文的研究结果和进一步的讨论。

## 2 文献综述

有些观测结果与其他观测结果相差太大,让人怀疑它们是由不同的机制产生的,这种现象被称为异常,有时也称为离群值[3, 9, 10]

。 多年来,异常检测一直被研究。
根据在训练过程中是否使用标签,异常检测可以分为三类 :监督异常检测、半监督异常检测和无监督异常检测[3, 4]

。

寻找与预期行为大相径庭的潜在时间序列数据模式被称为异常检测。在他们的研究[11]中,时间卷积网络框架被用作

预测模型,并采用多元高斯分布来识别时间序列中的异常点。

这种解决时间序列异常检测问题的方法称为无监督学习。

刘等[4]提出了一种基于时间卷积网络和高斯混合模型的时间序列在线异常检测框架。

该框架可以将高维时间序列映射到低维特征

空间,有利于有效的异常检测。实验结果表明,优化的时间卷积网络能够从时间序列数据中提取显着且有判别性的特征。

此外,高斯混合模型 (GMM) 在检测异常方面表现出很强的泛化能力和可靠性。具有贝叶斯推理的 GMM 能够有效检测异常,保持较低的误报率 (FPR) 0.796%,同时不影响 99.67% 的高召回率。其适用性扩展到时间序列异常检测。

Lorgat 等人 [12]提出了网络流量数据异常检测的通用框架,作为一种具有四个显著特征的时间序列数据 :自相似性、长程依赖性、非高斯性、

和非平稳性。

他和赵[11]实验了时间卷积网络（TCN)进行异常检测。

他们表明时间卷积网络可以自动学习序列数据中的固有模式。TCN 在不同的真实世界数据集、心电图 (ECG)、航天飞机和 2D 手势上进行了测试。基于 TCN 的方法在三个数据集上效果良好。结果显示 ECG 数据集的精度和 F1 分数更高,分别为 0.930 和 0.901。

比较生成的数据和原始数据之间的差异是基于 VAE 的方法的基本检测前提。对于测试数据,如果差异很小,则测试数据和训练数据很可能属于同一类,这表明测试数据是正常的。如果差异很大,则测试数据可能属于训练数据的相反类;在这种情况下,测试数据可能是异常数据[13]

。

An 和Cho[9]利用从 VAE 获得的重构概率提出了一种检测异常的方法。通过考虑可变性的概念,重构概率结合了 VAE 的概率特征。根据实验结果,该方法优于基于自动编码器和基于主成分的算法。通过利用变分自动编码器的生成特性,可以重建数据并了解异常的根本原因。

到

张等人[13]提出了一种基于概率质量函数（PMF)的新型特征编码技术。

这种编码技术有助于生成模型处理分类特征。他们还提出了一种基于对抗学习推理 (ALI) 的新检测方法。实验结果表明,他们的方法具有检测准确率高、训练效率高的优点,并表现出识别不熟悉攻击的能力。表 1 显示了他们的方法与基于二分类的支持向量机 (SVM)、标准一类 SVM (OC-SVM)、预先选择 3 个特征的一类 SVM (OC-SVM-3) 和基于生成对抗网络 (GAN) 的方法的比较。

评价指标为检测准确率F1-measure,值越大表示准确率越高;第二个评价指标为训练集

表1 五种方法的实验结果。

| 方法 | F1-measure 训练时间 (s) 测试时间 (s) 37.9 | | |
|---|---|---|---|
| 支持向量机 | 0.8318 | | 1.4 |
| OC支持向量机 | 0.0078 | 940.5 | 257.5 |
| OC-SVM-3 | 0.9691 | 9.9 | 1.8 |
| 然而 | 0.9247 | 1223.0 | 7291.0 |
| 但 | 0.9602 | 980.3 | 6.8 |

时间和测试时间,时间越短表示效率越高。表1显示他们的方法（ALI)在F1测量和效率方面均优于GAN。

OC-SVM-3 表现出最佳性能和效率,但这种方法涉及手动特征预选择。

Islam 等人 [8]尝试克服传统合成过采样技术的缺点,他们提出使用 VAE 来增强碰撞数据。使用 VAE 在数据生成过程中准确选择决策边界的能力有助于降低合成过采样技术中存在的过度拟合。

Ahn 等人[14]评估了各种基于深度学习的算法,用于发现航天器姿态控制系统中的异常。该研究旨在证明基于所选神经网络模型的异常检测方法的可行性，

评估航天器姿态控制系统的状况。

Kingma 和 Welling[7]提出了一种新型深度神经网络 VAE,用于顺序数据编码和解码。该模型提高了简单 GMM 方法在编码输入向量时的离散性。

Pangione 等人 [15]提出了一种在用少量标准数据进行训练时检测机器人异常数据的技术。他们在研究中采用了 VAE 来发现机器人手套箱配置中的异常。测量信号与机器人健康状况之间关系极其复杂且结构化,这是这一决定背后的驱动力。

Guo 等人 [3]发表了一种基于 GRU 的高斯混合 VAE,用于无监督异常检测。论文表明,只需在架构中添加时间数据处理能力,就可以使 VAE 模型从数据中识别出长期依赖关系。

Ahn 等人 [14]使用深度生成模型来检测异常并表征航天器姿态控制

系统故障。他们发现,即使模型只在正常数据上训练,VAE 也可以生成异常样本,而不是正常样本。他们将这一观察结果归因于对正常和

由于 VAE 模型的重参数化层采样后重建过程中产生的噪声而导致的数据异常。

自 Kingma 和 Welling 首次提出 [7] 以来,已有许多研究专注于展示 VAE 异常检测用法或使用时间特征进行调整。与 TCN 或基于 GRU 的高斯混合 AVE 方法不同,我们的研究从非时间角度辨别异常检测,因为并非所有异常检测问题都可以构造为时间序列。此外,我们认识到 VAE 重建过程中产生的噪声可能会扰乱正常数据和异常数据的分离。因此,我们提出了一种重新缩放 VAE 潜在空间的方法,并评估其对性能的影响。

# 3 研究方法

## 3.1 数据描述

本研究中使用的数据集取自开源项目和数据集共享网站 kaggle.com。该数据集是在 worldline 和比利时布鲁塞尔自由大学 (ULB) 机器学习小组 (http://mlg.ulb.ac.be) 的大数据挖掘和欺诈检测研究合作期间收集和分析的。该数据集包含欧洲持卡人在 2013 年 9 月使用信用卡进行的交易。该数据集包含 284 807 笔交易和 30 个数字特征。该数据集高度不平衡。正类 (492

欺诈交易（英语:frauding trading)占所有交易的0.172%。出于保密性考虑,数据集经过主成分分析（PCA)变换,V1至V28的特征被命名为PCA变换后得到的主成分。除了这些主成分之外，“时间”和“金额”特征没有进行变换。

## 3.2 研究框架概述

图 1 展示了 VAE 训练阶段的框架概览。在这个阶段,我们首先从数据集中过滤特征,只关注最具判别性的特征,然后从原始数据集中随机抽取正常和异常训练集。

正常与异常样本量严重不平衡。然后我们训练两个 VAE 模型。第一个

模型（VAE-normal)仅使用正常样本进行训练,另一个模型（VAE-anomaly)仅使用异常样本进行训练。特征选择的更多细节将在第 3.3 节中介绍。

图 2 展示了在同一测试集上测试两个 VAE 模型的框架的评估阶段

具有平衡类别比率的数据集。在评估中,我们首先通过对基于 F1 分数优化的重构误差进行阈值化来评估两个 VAE 模型的分类性能。然后,我们通过重构误差重新调整潜在空间以观察对分类性能的影响。第 3.4 节对数据采样和模型构建提供了更多解释。

在图 1 和图 2 中,潜在空间表示为L,重新缩放的潜在空间表示为Lnew。对于 VAE-normal,我们将模型重构误差表示为E

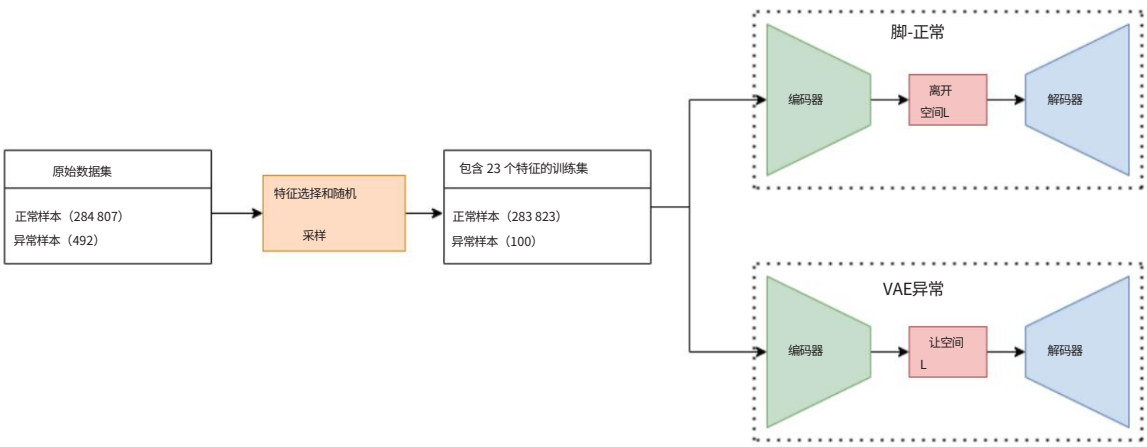



图1 VAE 模型训练概览。

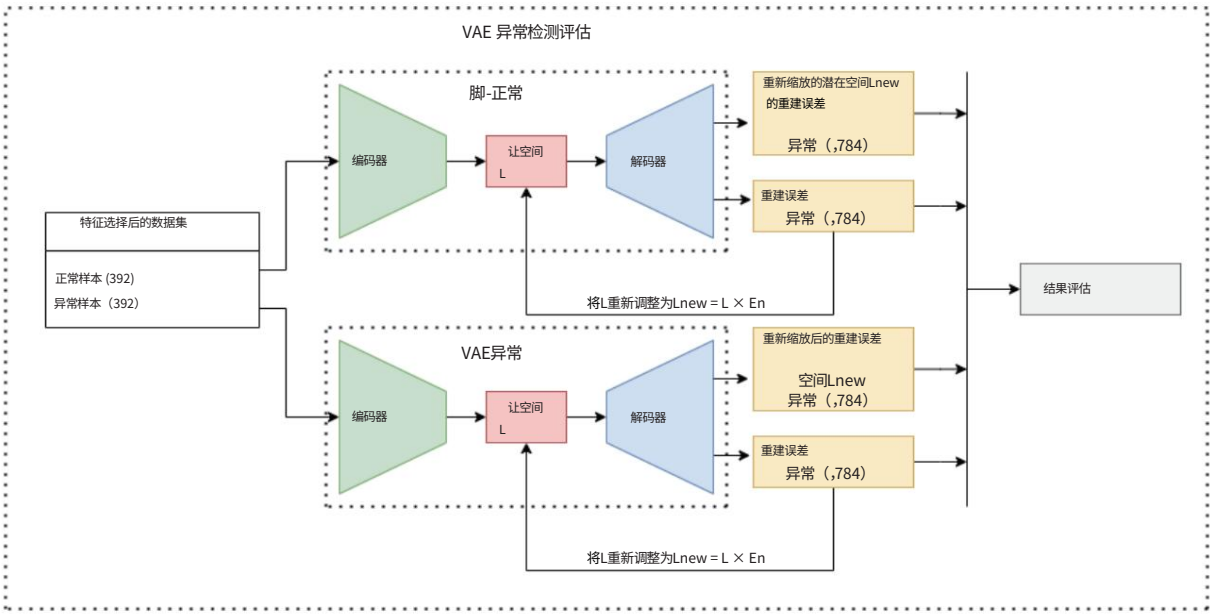



图 2 VAE 模型评估和潜在空间重新缩放概述。

和 VAE-anomaly 模型中,这些重建误差分别表示为Enormal和Eanomaly 。我们还将缩放系数表示为n。因此,潜在重缩放表示为Lnew = L × En,其中n是重缩放系数。

,

3.3 数据探索与特征选择

在训练模型之前,有必要确认正常数据和异常数据之间有足够的区别。需要检查两个关键因素,即特征共线性检验和双侧 Kolmogorov-Smirnov (KS) 检验。共线性检验探讨了两个类别的特征之间的关系。这有助于确定特征对在异常数据中的表现是否不同，从而为特征工程提供理论依据。双侧 Kolmogorov-Smirnov 检验是一种统计检验,用于确定两个底层一维概率分布是否不同。本研究将正常数据集中的每个特征与异常数据集进行测试。

我们绘制了正常和异常类别的两个相关热图来检查共线性。图 3 显示了两个类别之间的热图。

图 3 显示,正常类别的特征之间始终具有较低的负相关性,而异常类别则表现出明显较高的正相关性,变化也更显著。它还显示了 V1 到 V28 和金额特征。深蓝色（1.00）表示强正相关性，

深黄色表示强负相关（-1.00）。V1至V18表示异常类内最显著的变化。在这一步,可以确认这两个类具有明显的特征共线性和一些判别性特征。

第二项测试是双侧 KS 测试,用于比较正常类别和异常类别之间的特征。

图 4 显示了结果。

由于 VAE 模型试图近似样本的分布,因此检查特征分布是否不同也很重要

在异常类中。这可以通过可视化分布图和双侧 Kolmogorov-Smirnov 检验来实现。由于正常数据明显大于异常数据,因此对从正常数据和所有 492 个异常数据中随机抽取的 492 个正常样本进行双侧 Kolmogorov-Smirnov 检验。图 4 中的可视化显示了每个特征在正常和异常类中的分布

异常类。大多数特征的 Kolmogorov-Smirnov 值几乎为零,无法拒绝正态分布和异常分布相同的零假设。然而,特征 V13、V15、V22、V24、V25 和 V26 显示出明显更大的 Kolmogorov-Smirnov 值,并且它们的分布非常接近。这些特征在图 4 中以红色突出显示。

根据这一探索性分析,可以发现异常类别与

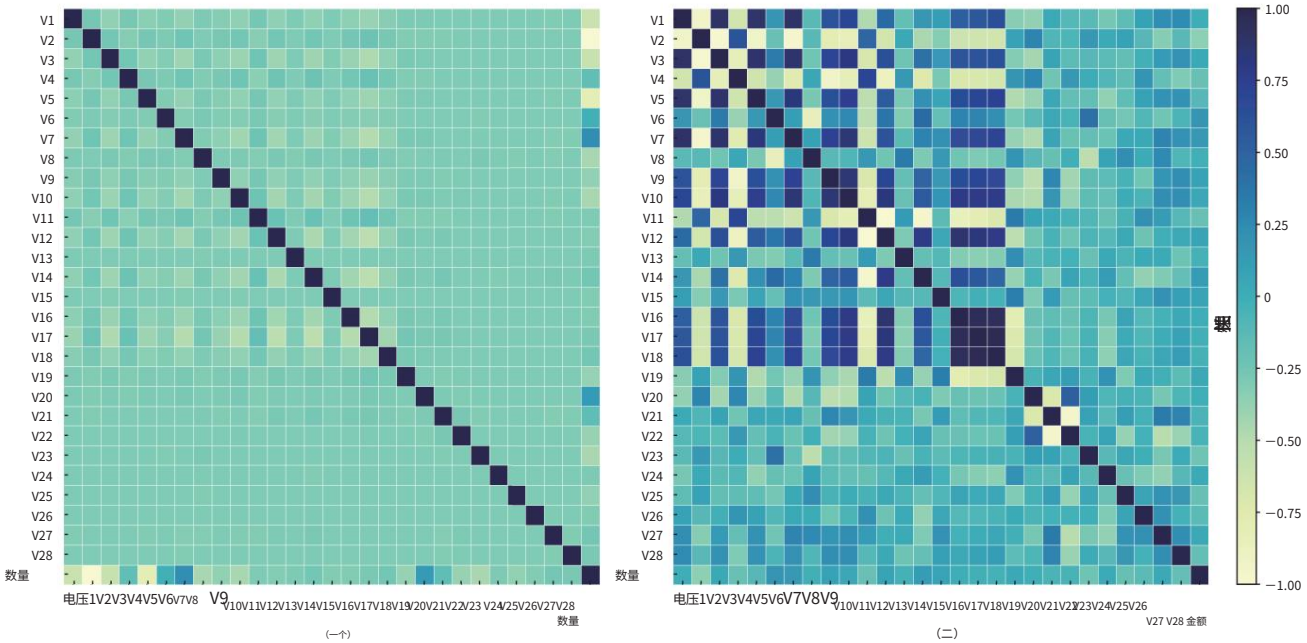



图 3 (a) 正常类别和 (b) 异常类别的热图。

正常类。此外,由于 V13、V15、V22、V24、V25 和 V26 的分布相似,这些特征会从训练和测试数据集中删除,因为它们对 VAE 模型来说是噪声。

总共选择了 23 个特征,包括 V1 到 V12、V14、V16 到 V21、V23、V27、V28 和数量。

3.4 模型构建与实验方法

与其他生成模型一样,VAE 使用潜在变量,通过近似联合概率分布来模拟输入数据分布[7]

。这种近似是通过使用神经网络实现的。

卷积神经网络 (CNN) 是一种流行的深度学习算法,主要用于计算机视觉和图像处理。顾名思义,基于 CNN 的模型使用方形滤波器矩阵在每一层执行离散卷积运算。在计算机视觉和图像处理中,CNN 已被证明能够捕捉图像中的空间和时间依赖性[11] 。

在我们的案例中,我们选择 CNN 作为 VAE 模型的基础,这样就可以捕获正常数据和异常数据的高级特征并将其发送到模型的下一级,最终反映在潜在空间中。此外,使用 CNN 需要更少的数据预处理,并且由于每层中的卷积运算,总训练参数会减少。在本研究中,我们评估了两个类似的模型。

其中一个 VAE 模型是在正常数据上训练的

第一个模型是在正常数据上训练的(VAE-normal),第二个模型是在异常数据上训练的(VAE-anomalous)。图 5 展示了两个 VAE 模型的编码器和解码器的结构。

我们从 492 笔异常交易中抽取 100 笔异常数据。然后,我们抽取 392 笔正常数据,并将其与其余 392 笔数据合并

异常数据。这个平衡比率数据集(392 个正常数据和 392 个异常数据)被保留用于两个模型的评估。第一个 VAE 模型(VAE-normal)在其余正常数据集(283 823 个样本)上进行训练,第二个 VAE 模型(VAE-anomaly)在 100 个异常数据上进行训练。

VAE 损失函数包括

重建误差和 KL-Divergence 损失,我们仅将重建误差用于异常检测任务。异常检测部分使用的重建误差是输入和重建向量之间的均方根误差 (RMSE),

$$\text{均方根误差} = \sqrt{\frac{\sum (v1 - v2)^2}{2}}$$

,其中 是输入向量,

v2 是重建的向量。为了检测异常,我们将训练和测试数据集传递给经过训练的 VAE 模型,并收到两个重建列表

误差。例如,如果 VAE-normal 仅在正常数据集上训练,则将异常数据传递给模型会导致更高的重构误差。然后,我们通过将重构误差设置为

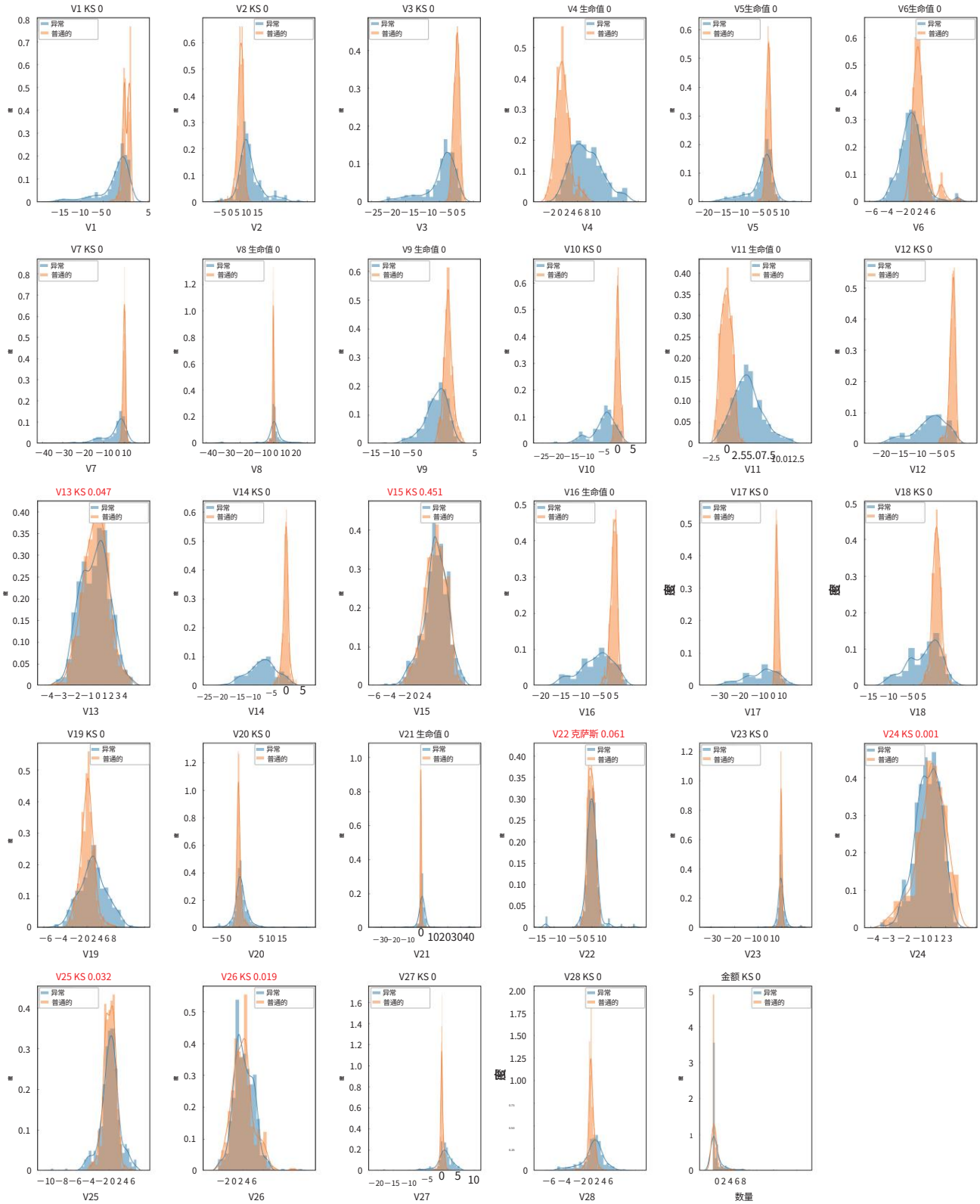
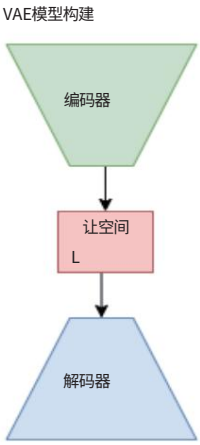



图 4 正常类和异常类之间的特征分布可视化,红色特征表示高柯尔莫哥洛夫-斯米尔诺夫值。

在训练数据集重建中选择四分位数
错误列表。然后我们在列表中使用此阈值

测试数据集重建误差并评估
F1 -score。导致最高分数的阈值




| 图层（类型） | 输出形状 | 数量参数 |
|---|---|---|
| 编码器输入（输入层） | [(无, 23, 1)] | 0 |
| 编码器_conv_0 (Conw1D) | （无,23,16） | 96 |
| 编码器_conv_1 (Conv1D) | （无,12,32） | 2592 |
| 编码器_conv_2 (Conv1D) | （无,12,64） | 10304 |
| flatten_41（展平） | （无,768） | 8 |
| 密集_85（密集） | （无,32） | 24,608 |
| Latent_in（密集） | （无,3） | 99 |
| Latent_log_var（密集） | （无,3） | 99 |
| samples_41（采样） | [(无,3),(无,3)] | 0 |
| 解码器_输入（输入层） | [(无,3)] | 0 |
| 密集_87（密集） | （无,768） | 3072 |
| reshape_45（重塑） | （无,12,64） | 0 |
| 解码器_conv_0 (Conv1DTransp) | （无,12,64） | 20,544 |
| 解码器_conv_1 (Conv1DTransp) | （无,24,32） | 10 272 |
| 解码器_conv_2 (Conv1DTransp) | （无,23,1） | 161 |
| cropping1d_41 (裁剪1D) | （无,23,1） | 0 |

图5 基于CNN的VAE模型的编码器和解码器结构。

然后选择 F1 分数。F1分数计算如下

$$\text{F1 分数} = \frac{\text{城市}}{\text{TP+2} \quad \frac{1}{-} \quad (\text{FP+FN})},$$

在哪里 城市 为真阳性率， FN阳性率,为假阴性率。 计划生育 是假的

数据集中的欺诈交易为“真”类,在模型预测中为负数。相反，“异常”类则是数据集中的欺诈交易,在模型预测中为正数。在混淆矩阵中,真负数是模型将真实非欺诈样本的数量预测为非欺诈,真正数是模型将真实欺诈样本的数量预测为欺诈,假负数是模型将真实欺诈样本的数量预测为欺诈,假正数是模型将真实非欺诈样本的数量预测为欺诈。

我们还研究了重新调整潜在空间对模型性能的影响。我们通过重构误差重新调整潜在分布： Lnew = L× E在相同的异常检测过程中,我们使用Lnew

n 。

针对 VAE-normal 和 VAE- 均进行了上述说明

异常。然后,我们将使用Lnew L的结果与使用原始的结果进行比较。

## 4 数据分析与结果

### 4.1 模型性能和潜在空间可视化

我们通过模型性能和潜在空间鲁棒性来评估模型。为了评估模型性能,我们使用混淆矩阵和F1分数。为了评估潜在空间,我们可视化编码训练输入和测试输入的潜在分布。图 6 显示了结果。在分类报告中，“正常”类别是非

我们的模型总体F1 得分为 0.92，

假阴性率略高。由于 VAE 的潜在空间是连续的,正常样本和异常样本的潜在表示不能线性或接近线性地分离。因此,当潜在空间重叠区域中的样本很可能被错误分类时。在我们的异常检测案例中,特定的异常样本与正常样本非常相似。下面的 3D 潜在空间可视化进一步说明了这一点。

在图 7 中,绿色组为潜在空间中的正常测试样本,红色组为异常测试样本。请注意，绿色组

| 班级 | 精确召回率 | | F1 分数 | 总样本数 |
|---|---|---|---|---|
| 正常（非欺诈） | 0.89 | 0.96 | 0.92 | 392 |
| 异常（欺诈） | 0.96 | 0.88 | 0.91 | 392 |
| 准确性 | – | – | 0.92 | 784 |
| 宏观平均值 | 0.92 | 0.92 | 0.92 | 784 |
| 加权平均值 | 0.92 | 0.92 | 0.92 | 784 |

| 班级 | 精确召回率 | | F1 分数 | 总样本数 |
|---|---|---|---|---|
| 正常（非欺诈） | 0.89 | 0.96 | 0.92 | 392 |
| 异常（欺诈） | 0.96 | 0.88 | 0.92 | 392 |
| 准确性 | – | – | 0.92 | 784 |
| 宏观平均值 | 0.93 | 0.92 | 0.92 | 784 |
| 加权平均值 | 0.93 | 0.92 | 0.92 | 784 |

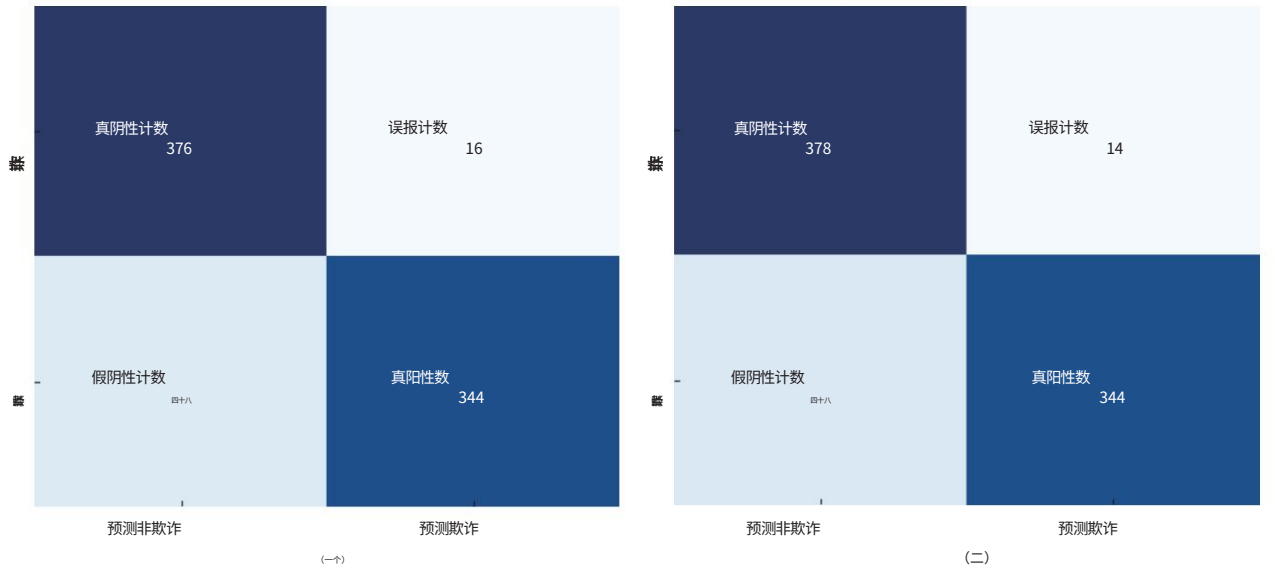



图 6 潜在空间重新缩放之前（a)和之后（b)的 VAE-normal 分类报告和混淆矩阵。

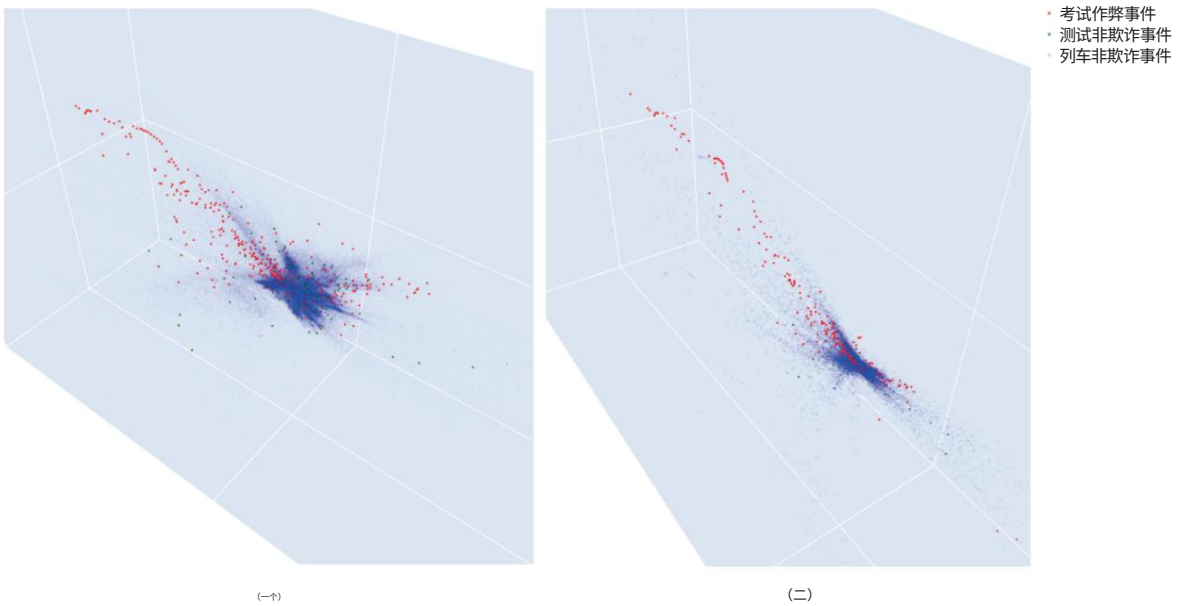



图 7 重新缩放前 (a) 和重新缩放后 (b) 的 VAE-normal 潜在空间。

与正常训练样本重叠
潜在空间的“中心”。这意味着我们基于 CNN 的 VAE
模型可以近似训练数据
分布作为正态分布的潜在表示
从重叠集群中进行训练和测试。在
另一方面,红色组也形成了一个集群
潜在空间,这意味着我们的模型也捕捉到了
异常类别的类似特征。

另一项评估涉及测试我们的模型
通过使用
重建误差向量。对于训练良好的模型，
这意味着潜在空间的中心变成
由于模型是在正常样本上训练的,因此密度更高，
这些样本的重建误差很低。因此，

大多数正常样本将被拉向
潜在空间的中心。相反,异常
在潜在空间中测试样本将变得更加
由于重建误差较大而扩散。这
缩放可以更好地分离潜在空间并暴露
容易被错误分类的样本;但是,如果
模型训练不正确或无法
近似生成训练集的分布
数据,缩放潜在空间以使其更加均匀
散布在中心周围并失去聚类
特征与缩放前相同。

图 7 显示了潜在空间缩放的结果
其中n = 1 。测试数据集中的正常样本
(绿色)明显集中在
潜在空间的质心(蓝色
组)。此外,来自
测试数据集(红色)在潜在
空间而不会失去其簇形状。然后我们使用
重新调整潜在空间以用于异常检测。
图 7 显示了错误
重新调整潜在空间后的阳性率。
预计,这是由于
潜在空间中的正常和异常数据
重新缩放。我们观察到与 VAE 异常模型类似的结果。

如图 8 所示,VAE 异常模型的错误

| 班级 | 精确召回率 | | F1 分数 | 总样本数 |
|---|---|---|---|---|
| 正常(非欺诈) | 0.87 | 0.97 | 0.92 | 392 |
| 异常(欺诈) | 0.97 | 0.85 | 0.91 | 392 |
| 准确性 | – | – | 0.91 | 784 |
| 宏观平均值 | 0.92 | 0.91 | 0.91 | 784 |
| 加权平均值 | 0.92 | 0.91 | 0.91 | 784 |

| 班级 | 精确召回率 | | F1 分数 | 总样本数 |
|---|---|---|---|---|
| 正常(非欺诈) | 0.89 | 0.96 | 0.93 | 392 |
| 异常(欺诈) | 0.96 | 0.88 | 0.92 | 392 |
| 准确性 | – | – | 0.92 | 784 |
| 宏观平均值 | 0.93 | 0.92 | 0.92 | 784 |
| 加权平均值 | 0.93 | 0.92 | 0.92 | 784 |

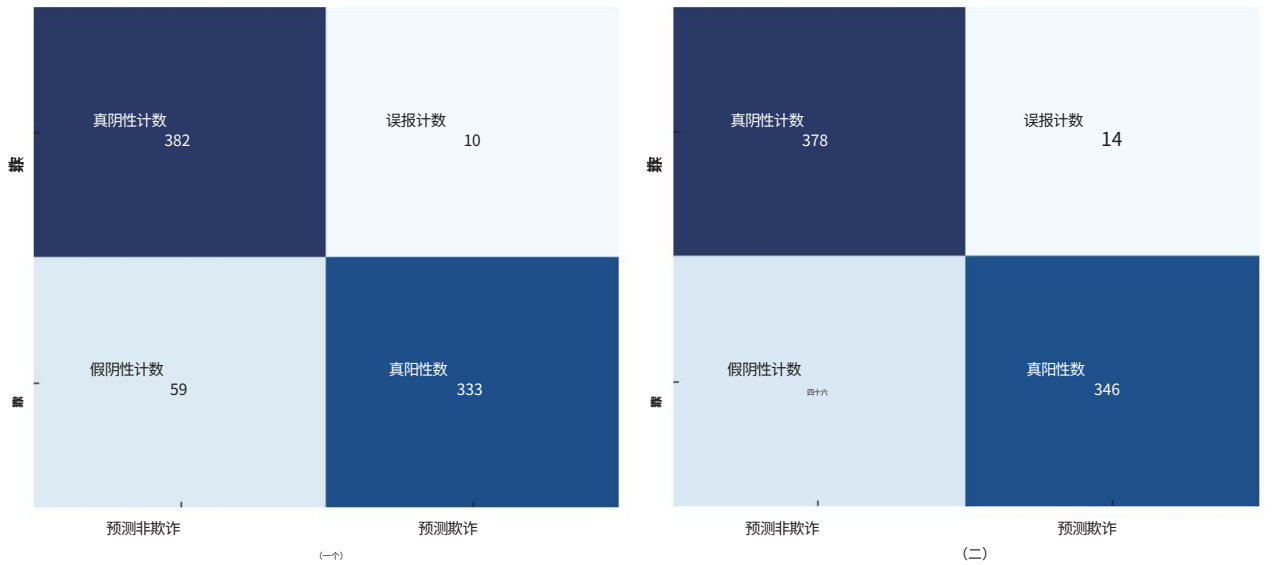



图 8 潜在空间重新缩放之前(a)和之后(b)的 VAE 异常分类报告和混淆矩阵。

阴性率提高,但假阳性率恶化。
重新调整后,总体F1得分增加了 1%
潜在空间与误差向量。与 VAE-normal 模型不同,VAE-anomaly
仅在 100 个样本上进行训练
异常样本,并在
小数据集往往拟合不足。然而,这
基于CNN的VAE模型仍然取得高分。 F1
此外,通过重新调整
潜在空间。图 9 显示了
重新缩放之前和之后的潜在空间。

4.2 与其他方法的比较

由于这项研究主要关注的是 VAE 在极端
不平衡数据集,我们将结果与其他结果进行比较
数据重采样实验和最终
性能指标。大多数实验都在这个数据集上进行
尝试首先解决数据不平衡问题。最近的
采样[17]。 包括尝试 过度采样[16, 17] , 在下面-
以及使用合成少数群体
过采样技术(即 SMOTE)
使用合成数据进行过度采样[18] 。表 2 显示
与其他方法的比较结果。我们重点介绍
我们的模型为“VAE-normal”、“VAE-anomaly”、“VAE-normal-rescaled”和“VAE-anomaly-rescaled”。

从表 2 可以看出,重新调整潜在
空间达到最高精度,XGBoost

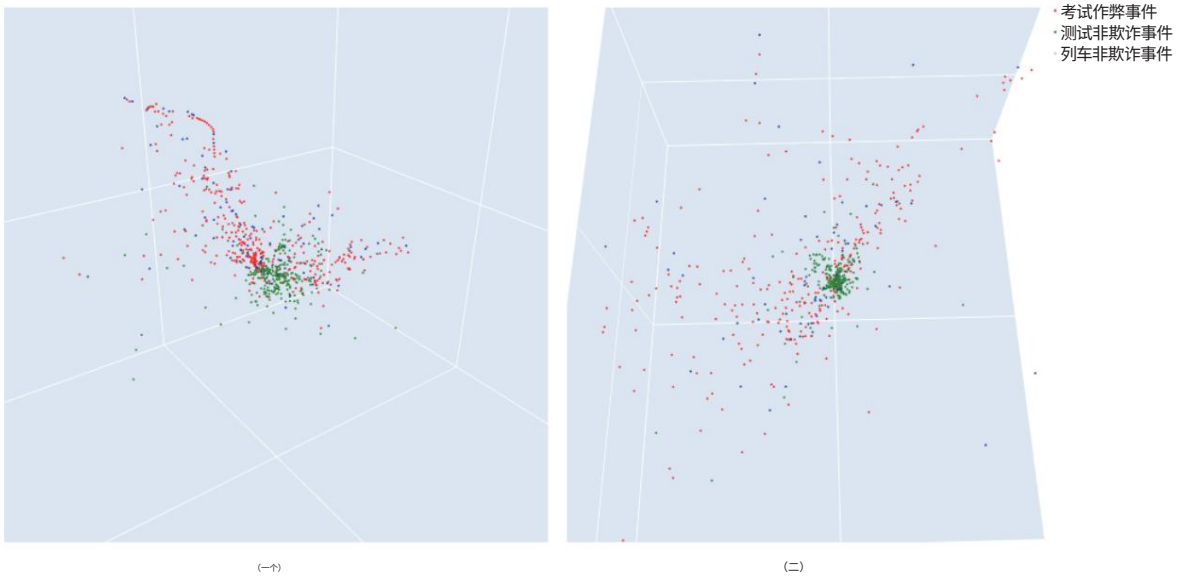



图 9 重新缩放前（a)和重新缩放后（b)的 VAE 异常潜在空间。

表2 与其他方法的比较结果。

| 模型 | 重采样方法精度 | | 记起 | F1 分数 |
|---|---|---|---|---|
| 脚-正常 | 没有任何 | 0.92 | 0.92 | 0.92 |
| VAE异常 | 没有任何 | 0.92 | 0.91 | 0.91 |
| VAE 正常-重新缩放 | 没有任何 | 0.92 | 0.92 | 0.92 |
| 重新调整 VAE 异常 | 没有任何 | 0.93 | 0.92 | 0.92 |
| 逻辑回归 | 欠采样 | 0.61 | 0.83 | 0.70 |
| 补充朴素贝叶斯 | 斯莫特 | — | — | 0.73 |
| 克尼恩 | 斯莫特 | — | — | 0.73 |
| 支持向量机 | 斯莫特 | — | — | 0.75 |
| 随机森林 | 斯莫特 | — | — | 0.75 |
| XGBoost | 欠采样 | 0.53 | 0.94 | 0.54 |
| XGBoost | 过度采样 | 0.70 | 0.93 | 0.77 |

欠采样实现了最高的召回率。

总体而言,本研究中的 VAE 模型实现了
最高的 F1 分数。请注意,VAE 模型实现了
这种性能不需要任何数据
增强或重采样方法。VAE-normal 和 VAE-anomaly 都实现了类似的结果
这表明即使
如果仅使用正常数据进行训练。

## 5 讨论与结论

在本研究中,我们构建了两个基于 CNN 的 VAE
模型、VAE-normal 和 VAE-anomaly,并评估
两种模型的潜在空间。我们还进行了实验
重新调整潜在空间
重构误差向量。VAE-normal 模型
未缩放和缩放后的 F1 分数均为 0.92
潜在空间。VAE 异常模型达到 0.90
未缩放潜在空间的 F1 分数和 0.92 F1 分数
具有缩放的潜在空间。我们观察到一些潜在的
潜在空间扩展的改进,尤其是当
训练数据不足,这可能会使
VAE模型比较适合分类任务。

使用基于 CCN 的 VAE 模型可以减少工作量
数据标记和数据预处理所需的。它
也绕过了大多数异常中的不平衡问题
检测场景。此外，
潜在空间提供了模型如何
学习和预测,从而使模型更加
可以解释且值得信赖。

在异常检测中使用 VAE 的一个限制是
分析连续潜伏的复杂性
空间。潜在空间提供了一个低维

表示输入,但不够专注于通过独特特征聚合它们。我们的研究发现,即使大多数异常输入聚集在潜在空间中,它们仍然与正常输入的中心簇分离得不够。我们发现重新调整潜在空间会对模型性能产生积极影响,尤其是在训练数据不足的情况下。

对于未来的研究,我们建议在潜在空间中测试一些基于聚类的采样方法,以便在复杂和非线性潜在空间中更好地分离低重构误差区域和高重构误差区域。

参考


[1] S. Zavrak 和 M. skefiyeli,使用变分自动编码器从网络流特征中进行基于异常的入侵检测, IEEE Access,第 8 卷,第 108346-108358 页,2020 年。

[2] B. Lebichot、F. Braun、O. Caelen 和 M. Saerens,《基于图的半监督信用卡欺诈检测系统》,《第五届国际复杂网络及其应用研讨会论文集》,意大利米兰,2016 年,第 721-733 页;Y. Guo、W. Liao、Q. Wang、L. Yu、T. Ji 和 P. Li,《多维时间序列异常检测:一种基于 gru 的高斯混合变分自动编码器方法》,《第十届亚洲机器学习研究会议论

[3] 文集》,中国北京,2018 年,第 97-112 页。

[4] J. Liu、H. Zhu、Y. Liu、H. Wu、Y. Lan 和 X. Zhang,《使用时间卷积网络和高斯混合模型进行时间序列异常检测》, 《物理学杂志:会议系列》,第 1187 卷,第 4 期,第 042111-042121 页,2019 年。

[5] GS Chadha、I. Islam、A. Schwung 和 SX Ding,《基于深度卷积聚类的时间序列异常检测》, 《传感器》,第 21 卷,第 16 期,第 5488 页,2021 年。

[6] M. Soleh、ER Djuwitaningrum、M. Ramli 和 M. Indriasari,基于单点交叉的特征工程策略,用于大数据分析中的欺诈检测,

Faleh Alshameri是计算机信息系统教授。他获得了美国乔治梅森大学的博士学位。他的研究兴趣包括文本挖掘、图像挖掘、数据科学和大数据分析。他在同行评审的期刊上发表了多篇研究论文,

国际会议。

J. Phys.: Conf. Ser.,第 1566 卷,第 1 期,第 012049 页,2020 年。

[7] DP Kingma 和 M. Welling,《变分自动编码器简介》, Found. Trends® Mach. Learn.,第 12 卷,第 4 期,第 307–392 页,2019 年。

[8] Z. Islam、M. Abdel-Aty、Q. Cai 和 J. Yuan,《使用变分自动编码器进行碰撞数据增强》, Accid. Anal.
翻译,卷。 151,p。 105950, 2021。

[9] J. An 和 S. Cho,使用重建概率的变分自动编码器异常检测,http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf,2023 年。

[10] Z. Niu、K. Yu 和 X. Wu,基于 LSTM 的 VAE-GAN 用于时间序列异常检测, Sensors,第 20 卷,第 13 期,第 3738 页,2020 年。

[11] Y. He 和 J. Zhao,《时间序列异常检测的时间卷积网络》, 《物理学杂志,会议系列》,第 1213 卷,第 4 期,第 042050-042056 页,2019 年。

[12] M. Lorgat、A. Baghai-Wadji 和 A. McDonald,《面向网络流量时间序列异常检测的通用框架》,第 16 届欧洲网络战和安全会议论文集,爱尔兰都柏林,2017 年,第 252-260 页。

[13] L. Zhang、W. Yang、H. Gan、M. Li、X. Wang 和 G. Liang,基于 PMF 编码和对抗性学习推理的异常检测,物理学杂志,会议系列,第 1187 卷,第 5 期,第 052037-052047 页,2019 年。

[14] H. Ahn、D. Jung 和 H.-L. Choi,《基于深度生成模型的航天器控制系统异常检测》, 《传感器》,第 20 卷,第 7 期,第 1991 页,2020 年。

[15] L. Pangione、G. Burroughes 和 R. Skilton,《变分自动编码器用于识别机器人中的异常数据》, 《机器人学》,第 10 卷,第 3 期,第 93 页,2021 年。

[16] E. Bilir,UnderSample 与 OverSample,https://www.kaggle. com/code/emirbilir/undersample-vs-oversample,2023 年。

[17] A. Wiratma,MIE213311-Tagas 1-逻辑和随机树-CC 欺诈,https://www.kaggle.com/code/arkalilang wiratma/mie213311-tas-1-logistic-random-tree-cc-fraud,2023。

[18] F. Alshameri 和 R. Xia,《信用卡欺诈检测:SMOTE 重采样和机器学习模型性能评估》, 《Int. J. Bus. Intell. Data Min.》,第 23 卷,第 1 期,第 1-13 页,2023 年。


夏冉是一名数据科学家和工程师。
他获得了美国玛丽蒙特大学信息技术硕士学位。他的研究重点是计算机视觉、自然语言处理和异常检测中的深度学习和可解释人工智能。他在

评审期刊和会议。